Delineamento Inteiramente Casualizado

Lucas Santana da Cunha
http://www.uel.br/pessoal/lscunha



Universidade Estadual de Londrina

22 de setembro de 2018
Londrina

Introdução

- O Delineamento Inteiramente Casualizado (DIC) ou Planejamento Completamente Aleatorizado é o mais simples de todos os delineamentos experimentais.
- Os experimentos instalados de acordo com este delineamento são chamados de experimentos inteiramente casualizados ou experimentos inteiramente ao acaso.
- São instalados em situação de **homogeneidade**, por isso, são muito usados em laboratórios, casas de vegetação, etc.

Características

- As principais **características** deste delineamento são:
  1. Utiliza apenas os princípios da **repetição** e da **casualização** (não utiliza o controle local);
  2. Os tratamentos são distribuídos nas parcelas de forma inteiramente casual, com números iguais ou diferentes de repetições para os tratamentos.

Vantagens

- As principais **vantagens** deste delineamento em relação aos outros são:
  1. É um delineamento bastante flexível, pois o número de tratamentos e de repetições depende apenas do número de parcelas disponíveis;
  2. O número de repetições pode variar de um tratamento para outro, embora o ideal seja utilizar o mesmo número de repetições para todos os tratamentos;
  3. A análise estatística é simples, mesmo quando o número de repetições por tratamento é variável;
  4. O número de graus de liberdade para estimar o erro experimental (que é dado pelo desvio padrão residual) é o maior possível.

Desvantagens

- As principais **desvantagens** deste delineamento em relação aos outros são:

  1. Exige homogeneidade total das condições experimentais;
  2. Pode nos conduzir a uma estimativa bastante alta para a variância residual, pois, não utilizando o controle local, todas as variações entre as unidades experimentais (exceto as devidas aos tratamentos) são consideradas como variação do acaso.

Tabulação dos dados

- Suponha que haja a tratamentos ou diferentes níveis de um único fator que se queira comparar. A resposta observada de cada dos a tratamentos é uma variável aleatória. Os dados seriam da forma:

Tratamentos	Observações				Totais	Médias
1	y_{11}	y_{12}	$\cdots$	y_{1b}	T_1	$\bar{y}_1$
2	y_{21}	y_{22}	$\cdots$	y_{2b}	T_2	$\bar{y}_2$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{ab}	T_a	$\bar{y}_a$

em que y_{ij} representa a j-ésima observação do i-ésimo tratamento;

$$T_i = \sum_{j=1}^{b} y_{ij} \qquad \bar{y}_i = \sum_{j=1}^{b} \frac{y_{ij}}{b} \qquad e \qquad \bar{y} = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{y_{ij}}{ab}.$$

Modelo estatístico

- O modelo de médias para o DIC é

$$y_{ij} = \mu_i + \epsilon_{ij}, \qquad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{cases} \tag{1}$$

- O modelo de efeitos para o DIC é:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \qquad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{cases} \tag{2}$$

em que,

- y_{ij} é o valor observado para a variável resposta obtido para o i-ésimo tratamento em sua j-ésima repetição;
  μ_i é a média de cada tratamento μ_i;
  μ é a média das médias de cada tratamento μ_i;
  τ_i é o efeito do tratamento i no valor observado y_{ij};
  ϵ_{ij} é o erro experimental associado ao valor observado y_{ij};

- Quando se instala um experimento no delineamento inteiramente casualizado, o objetivo é, em geral, testar a igualdade de médias dos a tratamentos. **As hipóteses estatíticas são:**

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \cdots = \mu_a \\ H_1 &: \mu_i \neq \mu_i', \quad i \neq i' \qquad \text{Pelo menos duas média se diferem} \end{aligned}$$

- Mas, como definimos,

$$\mu = \frac{\sum_{i=1}^{a} \mu_i}{a}$$

Uma forma equivalente de escrever as hipóteses anteriores é em termos dos efeitos dos tratamentos τ_i, que é:

$$\begin{aligned} H_0 &: \tau_1 = \tau_2 = \cdots = \tau_a = 0 \\ H_1 &: \tau_i \neq 0 \qquad \text{Pelo menos um tratamento} \end{aligned}$$

Do modelo estatístico

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \qquad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{cases} \tag{3}$$

Tem-se que os estimadores de mínimos quadrados para μ e τ_i, são dados por:

$$\begin{cases} \hat{\mu} = \bar{y}; \\ \hat{\tau}_i = \bar{y}_i - \bar{y}; \quad i = 1, 2, \ldots, a. \end{cases}$$

Somas de quadrados

- E assim, tem-se que a soma de quadrados total que corresponde à soma de quadrados dos desvios de todos os dados em relação à média é particionada da seguinte forma:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y})^2=\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_i+\bar{y}_i-\bar{y})^2$$

fazendo simplificações algébricas, tem-se

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y})^2=\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_i-\bar{y})^2+\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_i)^2$$

$$SQ_{total}=SQ_{trat}+SQ_{res}$$

$$SQ_{total} = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{(\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij})^2}{ab}$$

$$SQ_{trat} = \sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_i - \bar{y})^2 = \frac{\sum_{i=1}^{a} T_i^2}{b} - \frac{(\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij})^2}{ab}$$

em que T_i é o total do i-ésimo tratamento.

$$SQ_{res} = SQ_{total} - SQ_{trat}$$

ANAVA

Para verificarmos se a hipótese nula (H_0) é rejeitada ou não, completa-se o seguinte **Quadro da Análise de Variância:**

Tabela 1: Quadro da Análise de Variância.

CV	G.L.	S.Q.	Q.M.	F_{calc}	F_{tab}
Tratamentos (Entre)	a - 1	SQTrat	$\frac{SQTrat}{a-1}$	$\frac{QMTrat}{QMRes}$	$F_{(\alpha;GL_{Trat},GL_{Res})}$
Resíduo (Dentro)	a(b-1)	SQRes	$\frac{SQRes}{a(b-1)}$	-	-
Total	ab - 1	SQTotal	-	-	-

Se $F_{cal} > F_{(\alpha;GL_{Trat},GL_{Res})}$, então **rejeita-se** H_0, ou seja, há pelo menos duas médias de tratamentos que diferem entre si.

Valor esperado QM

- Temos que o $QMTrat$ é um estimador não viesado de σ^2 se H_0 for verdadeira. Entretanto, se H_1 for verdadeira, então $QMTrat$ é um estimador viesado de σ^2, pois o valor esperado é

$$E(QMTrat) = \sigma^2 + n\sum_{i=1}^{a}\tau_i^2$$

$$E(QMRes) = \sigma^2$$

- Observe que $QMRes$ é um estimador não viesado de σ^2 independente da hipótese H_0 ser verdadeira ou falsa.

Pressuposições

A análise de variância para testar essas hipóteses **só é válida** se forem satisfeitas as seguintes condições:

1. **aditividade:** os efeitos devem se somar (não há interação);
2. **independência:** os erros (ϵ_{ij}) devem ser independentes;
3. **normalidade:** os erros (ϵ_{ij}) devem possuir uma distribuição normal (teste Shapiro-Wilks e análise gráfica);
4. **homocedasticidade ou homogeneidade de variâncias:** os erros (ϵ_{ij}) devem possuir uma variância comum σ^2 (teste de Bartlett, teste F máximo e teste de Levene);

Coeficiente de Variação

- Uma medida para avaliação da precisão de experimentos é o **coeficiente de variação** (CV), dado por:

$$CV = \frac{\sqrt{QM_{res}}}{\bar{y}}.100$$

- Quanto menor o CV, mais preciso o experimento. A título de classificação, pode-se utilizar a seguinte tabela

CV	Precisão
$< 10\%$	Alta
10 a 20%	Médio
20 a 30%	Baixa
$> 30\%$	Muito Baixa

Exemplo 1

Num experimento de competição de linhagens de aves, com objetivo de se comparar a média de pesos (kg) aos 63 dias de idade, foram comparadas 5 linhagens.

Linhagens	Pesos médios das parcelas				
Cobb	1,73	1,75	1,70	1,73	1,79
Pilch	1,61	1,59	1,64	1,64	1,65
River	1,75	1,80	1,83	1,73	1,85
Ross	1,79	1,87	1,85	1,83	1,74
Rubbard	1,63	1,66	1,61	1,69	1,46

a) Formule as hipóteses relativas a este experimento;

b) Escreva o modelo estatístico do experimento;

c) Apresente o quadro da Análise de Variância e cite quais hipóteses devem ser assumidas para validade da mesma;

d) Interprete, estatisticamente, o resultado do teste F e comente o resultado, na prática.

e) Comente sobre a precisão do experimento.

Exercício 1

Um técnico de laboratório quer comparar a resistência à ruptura de três marcas de fio e, inicialmente, ele limita sua análise aos seguintes resultados (em quilogramas):

Fio 1	18,0	16,4	15,7	19,6	16,5	18,2
Fio 2	21,1	17,8	18,6	20,8	17,9	19,0
Fio 3	16,5	17,8	16,1	16,2	-	-

Considerando que as pressuposições da análise de variância foram satisfeitas, verifique, ao nível de 1% de significância, se há pelo menos uma diferença entre as marcas de fio.